An Analysis of CAD Modeling Procedure Data Collection Using Synchronous and Retrospective Think Aloud Techniques

Michael D. Johnson(✉) and Karl Ye

Texas A&M University, College Station, Texas, 77843, USA
mdjohnson@tamu.edu, Kaiba545@yahoo.com

Abstract. CAD is a critical tool for engineers in the 21st century. To improve CAD usage and education, methods for assessing and evaluating modeling procedures and decision making are necessary. To this end, two common verbal data collection methods are assessed for analyzing CAD modeling procedures. Stimulated recall and concurrent think aloud are compared to each other and screen capture video data. While the concurrent think aloud method seems to increase the necessary modeling time, the think aloud requirement does not affect the proportion of time spent on particular activities. A novel method of using Cohens Kappa with time usage data was implemented to compare the audio methods to screen capture video data. Neither audio method showed significant agreement with the video data when corrected for chance agreement. It is likely that both video and audio data are required to observe significant insights with respect to CAD modeling procedures and decisions. Drawbacks and benefits associated with alternative methods are also highlighted.

Keywords: Design: analysis and design methods · UX and usability: evaluation methods and techniques · Stimulated recall · Think aloud

1 Introduction

Computer-aided design (CAD) tools are at the nexus of the product commercialization process. This makes CAD modeling a critical skill for the modern engineer. Understanding how engineers model components can inform the design process as well as design education. CAD education is often viewed as lacking by practitioners who complain that engineers are entering the workforce unable to adequately translate design ideas into digital artifacts. This is often blamed on the focus of most CAD education on “cookbook” button pushing (declarative knowledge) as opposed to strategic design thinking (procedural knowledge) [1, 2]. What is needed is a way to capture the CAD modeling activities and the intent of those activities. Verbal data allows for these to be captured during CAD modeling; verbal data allows for significantly more content to be captured [3].

Verbal data can be a rich source of information for determining the processes associated with an activity; these data can either be concurrent (or synchronous) or

M. Kurosu (Ed.): HCI 2017, Part I, LNCS 10271, pp. 313–324, 2017.
DOI: 10.1007/978-3-319-58071-5_24

retrospective [3]. The concurrent data collection process is often termed *think-aloud* or *concurrent think aloud*. In this type of data collection process, participants provide a running commentary of their activities and thought processes. The retrospective data collection method is usually in the form of a *stimulated recall*. In stimulated recall, some form of media (i.e., photos, audio recordings, videos) is used to stimulate the participant's memory of their activity and inform their commentary on their thoughts and procedures. While both methods have been used to capture verbal data for a variety of activities, they have drawbacks. Think-aloud techniques may not work when participants are engaged in an activity that requires a "heavy cognitive load", they may stop talking; retrospective or stimulated recall techniques may be too general and lack the desired details [3].

Concurrent think aloud has been used in a wide array of situations to collect verbal data. This has included educational research examining student problem solving in Sudoku [4] as well as spatial ability problems [5]. The ability to capture this qualitative data improves the understanding of how people are solving problems. One aspect that is a concern in concurrent think aloud is the effect of the data collection on the activity in question. As noted above, when significant cognitive effort is needed, this method may not be effective [3]. One study examining the use of a disk utility tool found that those participants using think aloud actually performed better (faster and with less errors) [6]. One area where concurrent think aloud is widely used to collect thought processes is in design. Tolbert and Cardella [7] use video, screen capture, and audio data to examine how students think about a design process. Srinivasan and Chakrabarti [8] also use think aloud to capture the design process, and note that the verbal data collected allows them to assess a design's novelty. Mentzer, Becker and Sutton [9] use think aloud to compare the processes of students and experts in designing a playground. Kelley, Capobianco and Kaluf [10] use think aloud and coded data to examine how much time students spend on particular design activities. Think aloud has also been used to examine the effect that CAD tools have on the design process [11].

Retrospective data collection, or specifically stimulated recall, has also been widely used across numerous areas of study. Stimulated recall provides the ability to capture declarative knowledge while providing limited capabilities for procedural knowledge [12]. Ryan [12] notes that in the case of retrospective data collection, a stimulus should be used and that the time between the activity and the data collection should be limited to prevent "memory decay". Artzt and Armour-Thomas [13] use stimulated recall to capture declarative knowledge related to the solving of math problems; students are shown videos of themselves and asked to examine their metacognitive processes. Stimulated recall has also been used to assess students thoughts and feelings when solving physics problems [14]. Surgeons have been shown videos of themselves operating and asked to explain their decision making processes [15]. Video informed stimulated recall has also been used to evaluate the decision and actions of teachers [16] and counselors [17]. Stimulated recall can also be informed by photographs [18]; however, trying to use non-video stimulated recall has its limitations [19].

While both concurrent and retrospective data collection have their strengths, multiple methods are needed to provide richer data [4]. Trevors, Feyzi-Behnagh, Azevedo and Bouchet [20] use a combination of eye tracking data, concurrent, and retrospective data verbal data collection to examine the understanding of science concepts. Bruun and Stage [21] examined alternative methods to assess usability; they found that coaching (or engaging the participant in conversation) performed better than silent observation. Kuate, Soh Fotsing and Kenmeugne [22] use both concurrent and retrospective methods examining CAD modeling; they find less stated information regarding design intent in the retrospective case. This contrasts the expected result of retrospective methods providing more information regarding declarative knowledge which would include design intent. The current study also examines CAD modeling procedure; both concurrent and retrospective methods are used.

2 Methods

The data presented in this work was collected during a junior-level computer aided design course. Near the end of the semester, after most significant instruction in the CAD program (Creo Parametric 2.0) had been completed, students were assigned one of three alternative modeling situations. These included modeling a component from a drawing of that component, modeling the same component from a physical representation of the component, or modeling a component of the student's choosing that they had brought from home (Fig. 1). These alternative modeling tasks were part of a broader research project, but for the purposes of this work were useful in providing variability in the tasks being assessed.

2.1 Video Capture and Modeling Procedure Analysis

All CAD modeling activity was recorded with the screen capture software Camtasia. This usage data was examined using a continuous time log to assess what the participant was doing in increments of seconds. The modeling time was divided into five categories [23]: doing, searching, thinking, trial and error, and waiting. Doing was defined as the actual modeling of the component (i.e., using tools and features to create geometry). Thinking was any time there was a lack of cursor movement or panning and rotating without a clear purpose. Searching was defined as looking for particular items by clicking on menu items or icons. Trial and Error was the creation of geometry and then its complete deletion at a later time. Waiting (or regeneration) time was the user waiting for the completion of graphical rendering or some other computational process.

2.2 Concurrent Think Aloud Data Capture and Analysis

Concurrent think aloud data was captured for a subset of the overall modeling group.

This included 8 participants: 3 modeling using the drawing, 3 with the physical model, and 2 with items from home. Students were given up to 75 min to finish their modeling activities. Each participant was told that they would be providing a running commentary of their modeling activities. They were all primed with a simple task of putting together a 9-piece puzzle. They were asked to talk about what they were doing and how they were solving the puzzle. After they completed the puzzle task both the

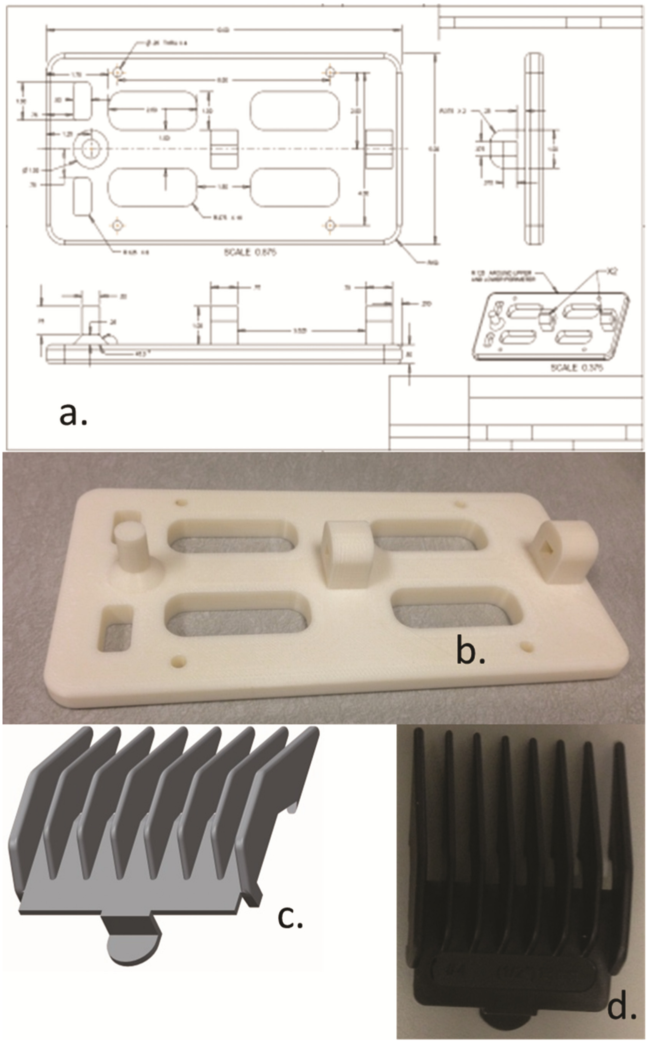

Fig. 1. Drawing of standard component (a), physical model (b), student selected component CAD model rendering (c), and photo of student selected component (d)

audio recording and the Camtasia screen capture program were started. The screen capture run time was noted on the audio recording to allow for the synchronization of the two data collection methods. The participants were prodded with the phrase “what are you doing now” whenever there was a period of silence. When the participant announced that they felt they were done, both the audio and screen capture recording were stopped.

The analysis of the audio data was similar to that of the screen capture data described above. However, given the inability to verbally differentiate between productive modeling (Doing) and modeling that was later deleted (and would therefore be coded as Trial and Error), Trial and Error was combined into doing. An additional category of indeterminate was also added capture audio data that could not be appropriately categorized. The time for each activity was noted in a running log in increments of seconds.

2.3 Stimulated Recall Data Capture and Analysis

Again, the stimulated recall data was collected from a subset of the overall modeling group. This included 9 participants: 4 modeling using the drawing, 2 with the physical model, and 3 with items from home. Again, students were given up to 75 min to complete their modeling activities. In the case of the stimulated recall data collection, screen capture data was collected without any interaction with the participant. Once the modeling was complete, participants were asked to watch the captured video of their modeling activities and comment on what they were doing at that point in the video. This discussion of their activity was audio recorded. Participants were prompted with the phrase "what are you doing now". Similar to the concurrent think aloud data these data were tabulated into the same categories as above (including an indeterminate one).

2.4 Comparison of Audio and Video Data

Cohen's [24] Kappa (κ) was used to determine the agreement between audio data and the screen capture video data (with combined Trial and Error and Doing). Indeterminate data was not included in the comparison of video and audio analysis since it has no video analog. Cohen's Kappa is often used to assess inter-rater agreement and takes into account chance agreement. In the case of this work, it is used to see if either the synchronous or asynchronous audio data provided better agreement with the video data. The amount of overlap in time for each category was tabulated. The agreement among the categories was corrected for chance to determine Cohen's Kappa. Readers interested in a more detailed description of this process are referred to Gwet [25]. Cohen's Kappa was also used to assess the agreement between two raters that analyzed the audio data. Two of the concurrent think aloud audio recordings were assessed and had an average κ value of 0.711. This represents substantial agreement [26]. An additional two audio recordings of the stimulated recall data were compared between two raters. These had an average κ value of 0.775.

3 Results

The goal of this work was to compare alternative audio data collection techniques for CAD modeling procedure analysis. The basis for the comparison is video data collected from the screen capture of the modeling activity. It should be noted that the audio data analysis and the screen capture data analysis were not done concurrently (i.e., the raters would not have recognized the audio of a video that they had previously viewed). Two raters were used; they were also not involved in the initial data collection process. While inter-rater agreement data is often reported for rating nominal data (e.g., a medical condition), this stringent condition is of use for methodological data as well. Often general agreement among raters or methods is reported; however this data does not take into account chance agreement. The use of Cohen's Kappa (κ) corrects for this chance agreement [24]. This allows for a quantitative basis to be used along with qualitative observations to evaluate the positive and negative aspects of the alternative audio data collection methods.

An example of the agreement between audio and video time usage data is shown in Table 1. In the case of the audio data, a doing proportion of 56.6% is reported; for the video data, it is 43.9%. When comparing the time logs for each of the data sources, an overlap of 39.2% total modeling time for doing was found. For the overall modeling activities, an overall percent agreement (P_0) of 71.5% was found. Given the alternative data sources, this would likely be seen as a high level of agreement. However, when correction for the chance agreement percentage (P_C) of 44.4% is taken into account, it is less impressive. This results in a Kappa (κ) of 0.488 from a maximum Kappa (κ_M) of 0.646. κ_M corrects for the disagreement of off diagonal results (i.e. the original distribution of time usage for each method).

Table 1. Video and audio data comparison analysis example

		Video				
		Doing	**Searching**	**Thinking**	**Waiting**	
Audio	**Doing**	39.2%				56.5%
	Searching		0.0%			5.7%
	Thinking			32.2%		35.4%
	Waiting				0.1%	0.5%
		43.9%	0.8%	55.1%	0.2%	

To examine if the data collection method (namely the think aloud requirement) had an effect on the modeling procedure the modeling times for each category and the overall modeling times were compared. These are shown in Table 2. All modeling time categories, along with the overall modeling time, were greater for the concurrent think aloud protocol. This could be due to the person having to provide a verbal commentary on what they are doing; this could slow them down or increase the cognitive load associated with their modeling activity. It should be noted that while the differences in the Doing and overall time categories are large (over 500 and 700 s, respectively), these differences are not statistically significant at the a = 0.05 level. There is also a chance that the small data set could result in individuals that are less skilled modelers taking longer and skewing the results. The composition of the modeling activities is such that the stimulated recall group should have an equal or greater time requirement; based on the authors' experience and observations, the individual items brought from home require the longest modeling time. There is a larger percentage of those items in the stimulate recall data set.

Table 2. Comparison of absolute time usage audio data for alternative methods

	Concurrent Think Aloud		Stimulated Recall			
	Mean	*SD*	Mean	*SD*	*t*	*p*
Doing (s)	2176.6	403.2	1672.3	571.5	2.075	0.056
Searching (s)	222.9	139.7	214.0	146.6	0.127	0.900
Thinking (s)	703.1	518.0	474.9	233.6	1.147	0.280
Waiting (s)	23.5	32.0	17.4	14.6	0.513	0.616
Total (s)	3148.3	851.8	2378.7	722.5	2.016	0.062

To correct for the effect of the think aloud requirement on the overall time required for the modeling activities, the percentages associated with each of the time usage categories were also compared. These are shown in Table 3. In this case, the results are comparable. Both the think aloud and the stimulated recall groups use approximately 70% of their modeling time for productive geometry creation in the Doing category. The next largest category of time usage was Thinking; it was approximately 20% in both cases. These results show that while the think aloud method may have extended the time, it did not alter the distribution of time that the participants used to complete certain tasks.

Table 3. Comparison of time usage audio percentage data for alternative methods

	Concurrent Think Aloud		Stimulated Recall			
	Mean	*SD*	Mean	*SD*	*t*	*p*
Doing (%)	71.3%	12.9%	69.9%	6.3%	0.276	0.788
Searching (%)	6.7%	2.6%	8.8%	5.3%	-1.002	0.332
Thinking (%)	20.5%	12.3%	20.4%	9.7%	0.023	0.982
Waiting (%)	0.7%	1.1%	0.8%	0.8%	-0.187	0.854

To compare the agreement between audio and video data for the two alternative audio data collection methods, their nominal agreement (P_0) along with their corrected Cohen's Kappa (κ) agreement were compared. The effect of collection method on chance agreement (P_C) maximum Kappa (κ_M) are also shown in Table 4. While the Kappa for stimulated recall is slightly higher than that of the think aloud method, this difference is not statistically significant. According to Landis and Koch [26] this agreement is deemed "slight" bordering on "fair". As noted above, the use of Cohen's Kappa is a stringent condition for this type of data. These results are less disappointing when examining the maximum Kappa (κ_M) data. Given the general disagreement between the audio and visual data sets, the best possible results would be deemed Moderate [26]. Even the nominal agreement (P_0) between the audio and visual data is only slightly greater than 50%. Overall there was no statistically significant difference between the agreement variables for the two audio data collection methods. The lack of significant agreement also does not allow one to say that either method would be preferable based on its agreement with the video data. The lack of agreement also does not allow for audio data to be a substitute for the video data analysis. The analysis of the screen capture videos typically required 3 to 4 times the modeling time for analysis (i.e., it required 4 h to analyze a 1 h video). The analysis time requirements for the audio data were closer to 2 times. As Vandevelde, Van Keer, Schellings and Van Hout-Wolters [4] pointed out, there is often a need for multiple methods.

Table 4. Comparison of audio and video agreement variables for alternative methods

	Concurrent Think Aloud		**Stimulated Recall**			
	Mean	***SD***	**Mean**	***SD***	***t***	***p***
$\boldsymbol{k}$	0.178	0.171	0.201	0.125	-0.321	0.752
$\boldsymbol{k_M}$	0.462	0.183	0.535	0.094	-1.016	0.333
$\boldsymbol{P_0}$	56.2%	10.4%	55.2%	3.9%	0.254	0.805
$\boldsymbol{P_C}$	46.7%	5.1%	43.2%	6.5%	1.244	0.233

The last aspect that was investigated was the relationship between Kappa and the various time usage categories. The percentages of the overall time usage was used for comparison given that it is not affected by the audio data collection method. These comparisons are shown in Table 5. Among the time usage variables, the percentage Doing time is negatively correlated with both the Searching time and the Waiting time. This negative correlations are statistically significant. This is an expected result; Searching is not productive use of time (like Thinking) and just adds to the overall modeling time. The same is true of Waiting. Kappa is negatively correlated with both Doing percentage and Thinking percentage; results are not statistically significant. Kappa is significantly positively correlated with the percentage Searching time Waiting time. This is an expected result as these two categories have less coding ambiguity with respect to video analysis. As the percentages of Searching and Waiting time increase, the ability to find agreement between the audio and video data increases. Overall, the results show that audio data collection is probably not a viable substitute for screen capture data analysis and that multiple methods are likely needed.

Table 5. Comparison of Cohen Kappa and activity categorization percentages

	2	3	4	5
1. Video – Percentage Doing	-0.632** (0.007)	0.057 (0.829)	-.690** (0.002)	-0.378 (0.134)
2. Video – Percentage Searching		-0.779** (0.000)	0.943** (0.000)	0.537* (0.026)
3. Video – Percentage Thinking			-0.752** (0.000)	-0.414 (0.099)
4. Video – Percentage Waiting				0.539* (0.026)
5. Kappa (k)				

Note: Significance shown in parentheses below correlation; *Correlation is significant at the 0.05 level (2-tailed); **Correlation is significant at the 0.01 level (2-tailed)

4 Discussion

Verbal or audio data is often collected to help understand thought processes and decision making. The two main methods for this data collection include a synchronous, concurrent think aloud, method as well as an asynchronous, stimulated recall, method. Both of these methods have been used to collect data around design processes and design decision making in general [8–10] and computer-aided design (CAD) in particular [11, 22]. This work compared these two verbal data collection methods to examine how participants used their time modeling various components in CAD. The time usage was tabulated into four main categories: Doing, Searching, Thinking, and Waiting. The tabulations for time usage from the audio data was compared to that collected from screen capture videos of the modeling process. Unique to this work, these comparisons used agreement between audio and video data based on running time logs and corrected for chance agreement using Cohen's Kappa [24].

Screen capture data analysis, concurrent think aloud, and stimulated recall all have their associated benefits and drawbacks. In the case of video analysis, the intent and activities of the participant must be inferred by the analyst without input from the participant. This is also a time consuming process that often involves pausing, rewinding and re-watching the video. Concurrent think aloud requires that the participant actively detail what they are doing while they engaged in the activity. If the activity is cognitively demanding, this may slow their verbal response or their progress in the activity. The concurrent think aloud data in this work was quicker to analyze than the screen capture video data and provided firsthand knowledge of what activities were being done. Concurrent think aloud methods may provide better results when procedural knowledge is sought. Stimulated recall requires that some stimulus be used to elicit a response for the activities under investigation; there are limits to the effectiveness of this method when a stimulus such as video is not available [19]. Stimulated recall has the ability to better capture declarative knowledge, but may not excel at capturing procedural knowledge [12]. Stimulated recall also puts an additional burden on the participant; they must engage in the activity and then relive the activity to provide the audio data. To assess these alternative methods, CAD modeling procedure data were collected using all three methodologies.

A comparison of the absolute time durations for the various activity categories showed that the concurrent think aloud modeling procedure took a longer time overall as well as for the various time usage categories than stimulated recall. This was an expected result; the requirement to verbally detail the procedures as well as carry them out added to the overall time. However, when comparing the proportional time usage tabulations, the two methods had very similar proportions for the various categories.

Unique to this work, a method for comparing the various time usage data sets and correcting for chance agreement was implemented. This allowed for the audio data collection methods to be compared to the video screen capture data. While often used for nominal comparisons between raters, Cohen's Kappa [24] was used to compare the time used for various modeling activities. While this is likely a more rigorous test of agreement than necessary for time usage data, it does provide for the necessary correction to account for chance agreement among the data. A comparison of the agreement variables did not show any significant differences for the two audio data collection methods when compared for agreement with the screen capture video data. The corrected agreement κ was also somewhat low for both methods; with κ averages of 0.178 for concurrent think aloud and 0.201 for stimulated recall, this agreement would be deemed slight or fair [26]. The correlations of κ with the various modeling proportions showed that agreement was significantly positively correlated with both Searching time and Waiting time. This is to be expected as these categories have less coding ambiguity with respect to agreement between audio and video analysis; increases in proportion of time spent doing these activities would increase agreement. Given the overall lack of agreement between the audio and video data, both are likely needed to provide quality insights into design processes and procedures.

4.1 Conclusions

CAD modeling procedure data for two verbal data collection methods, stimulated recall and concurrent think aloud were compared to each other and screen capture video data. While the concurrent think aloud method seemed to increase the necessary modeling time, it did not affect the proportion of time spent on particular activities. A novel method of using Cohens Kappa with time usage data was used to compare the audio methods to screen capture video data. Neither audio method showed significant agreement with the video data when corrected for chance agreement. Given this, a combination of both audio and video data are likely necessary to collect significant insights into modeling procedures and design decision making. Depending on the focus of the analysis either stimulated recall (for declarative knowledge focused work) or concurrent think aloud (for procedural knowledge focused work) can be used.

4.2 Limitations

The above conclusions should be viewed in light of the limitations associated with this work. First, the limited sample size of 8 for the concurrent think aloud method and 9 for the stimulated recall method limit its broad applicability. Also, it should be noted that the participants were all students. This may increase the variability of their modeling performance and affect the overall data set. Future work will attempt to increase the sample size and enlist professional CAD users to provide data.

Acknowledgements. This material is supported by the National Science Foundation under EEC Grant Number 1129403. Any opinions, findings, conclusions, or recommendations presented are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Hamade, R.F., Artail, H.A., Jaber, M.Y.: Evaluating the learning process of mechanical CAD students. Comput. Educ. **49**, 640–661 (2007)
2. Lang, G.T., Eberts, R.E., Gabel, M.G., Barash, M.M.: Extracting and using procedural knowledge in a CAD task. IEEE Trans. Eng. Manag. **38**, 257–268 (1991)

3. Ericsson, K.A., Simon, H.A.: Verbal reports as data. Psychol. Rev. **87**, 215–251 (1980)
4. Vandevelde, S., Van Keer, H., Schellings, G., Van Hout-Wolters, B.: Using think-aloud protocol analysis to gain in-depth insights into upper primary school children's self-regulated learning. Learn. Individ. Differ. **43**, 11–30 (2015)
5. Mohler, J.L.: Examining the spatial ability phenomenon from the student's perspective. Eng. Des. Graph. J. **72**, 1–15 (2008)
6. Wright, R.B., Converse, S.A.: Method bias and concurrent verbal protocol in software usability testing. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **36**, 1220–1224 (1992)
7. Tolbert, D., Cardella, M.E.: CAREER: mathematics as a gatekeeper to engineering: the interplay between mathematical thinking and design thinking - using video data. In: ASEE Annual Conference and Exposition, Conference Proceedings
8. Srinivasan, V., Chakrabarti, A.: Investigating novelty-outcome relationships in engineering design. Artif. Intell. Eng. Des. Anal. Manuf. AIEDAM **24**, 161–178 (2010)
9. Mentzer, N., Becker, K., Sutton, M.: Engineering design thinking: high school students' performance and knowledge. J. Eng. Educ. **104**, 417–432 (2015)
10. Kelley, T.R., Capobianco, B.M., Kaluf, K.J.: Concurrent think-aloud protocols to assess elementary design students. Int. J. Technol. Des. Educ. **25**, 521–540 (2015)
11. Salman, H.S., Laing, R., Conniff, A.: The impact of computer aided architectural design programs on conceptual design in an educational context. Des. Stud. **35**, 412–439 (2014)
12. Ryan, J.: Stimulated recall. In: Researching Language Teacher Cognition and Practice, pp. 144–161 (2012)
13. Artzt, A.F., Armour-Thomas, E.: Mathematical problem solving in small groups: exploring the interplay of students' metacognitive behaviors, perceptions, and ability levels. J. Math. Behav. **16**, 63–74 (1997)
14. Appleton, K.: Problem solving in science lessons: how students explore the problem space. Rese. Sci. Educ. **25**, 383–393 (1995)
15. Chen, X., Williams, R.G., Smink, D.S.: Dissecting attending surgeons' operating room guidance: factors that affect guidance decision making. J. Surg. Educ. **72**(6), e137–e144 (2015)
16. Hubber, P., Tytler, R., Haslam, F.: Teaching and learning about force with a representational focus: pedagogy and teacher change. Res. Sci. Educ. **40**, 5–28 (2010)
17. Stockton, R., Morran, D.K., Clark, M.B.: An investigation of group leaders' intentions. Group Dyn. **8**, 196–206 (2004)
18. Fox-Turnbull, W.H.: The nature of primary students' conversation in technology education. Int. J. Technol. Des. Educ. **26**(1), 21–41 (2015)
19. de Smet, M., van Keer, H., de Wever, B., Valcke, M.: Studying thought processes of online peer tutors through stimulated-recall interviews. High. Educ. **59**, 645–661 (2010)
20. Trevors, G., Feyzi-Behnagh, R., Azevedo, R., Bouchet, F.: Self-regulated learning processes vary as a function of epistemic beliefs and contexts: mixed method evidence from eye tracking and concurrent and retrospective reports. Learn. Instr. **42**, 31–46 (2016)
21. Bruun, A., Stage, J.: An empirical study of the effects of three think-aloud protocols on identification of usability problems. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) INTERACT 2015. LNCS, vol. 9297, pp. 159–176. Springer, Cham (2015). doi:10.1007/978-3-319-22668-2_14
22. Kuate, G., Soh Fotsing, B.D., Kenmeugne, B.: Restitution of design intents by computer aided design models. Int. J. Appl. Eng. Res. **7**, 277–291 (2012)

23. Johnson, M.D., Ozturk, E., Yalvac, B., Valverde, L., Peng, X., Liu, K.: A methodology for examining the role of adaptive expertise on CAD modeling. In: ASME 2015 International Mechanical Engineering Congress & Exposition, IMECE2015-50296. ASME, Houston (2015)
24. Cohen, J.: A coefficient of agreement for nominal scales. Educ. Psychol. Meas. **20**, 37–46 (1960)
25. Gwet, K.L.: Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters, 3rd edn. Advanced Analytics, LLC, Gaithersburg (2012)
26. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics **33**, 159–174 (1977)